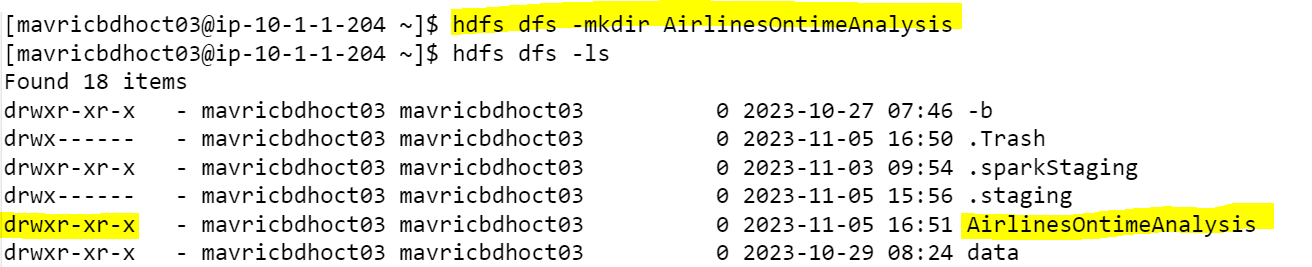
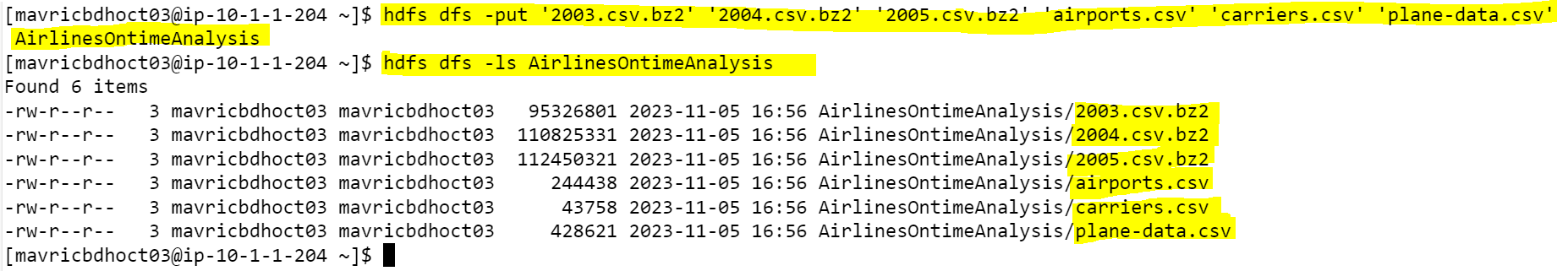
**OntimeFlightPerformance**

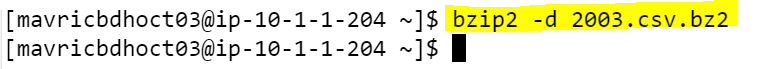
1. **Data Setup in HDFS**

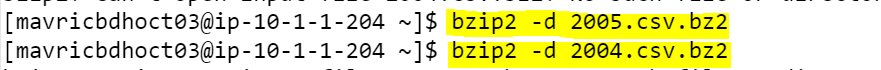
**Step 1: Copy the provided csv files in a directory named AirlinesOntimeAnalysis**

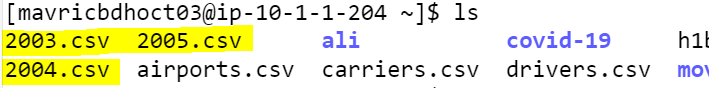




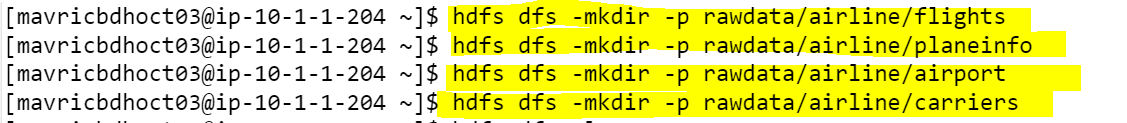
**Step 2: Unzip the contents for the bz2 files.**







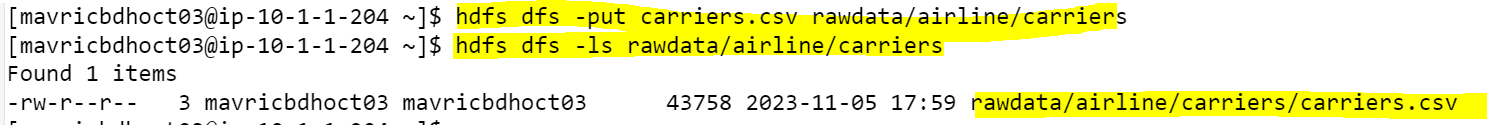
**Step 3: Create the directories in hdfs where your data will reside.**

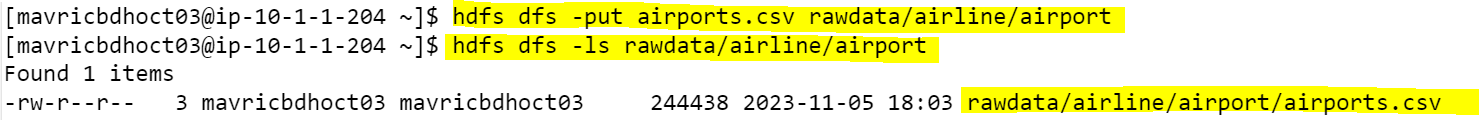


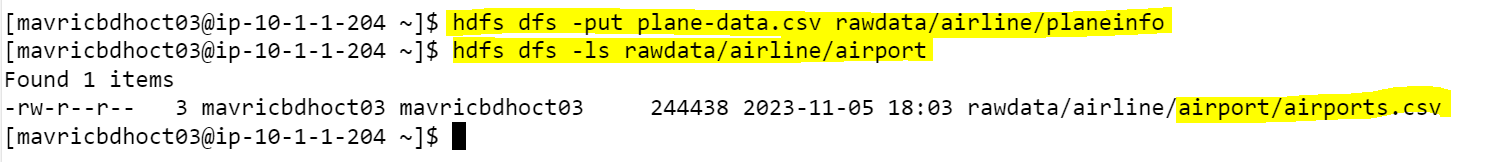
A screenshot of a computer

Description automatically generated

**Step 4: copy the respective files to the above created hdfs structure.**

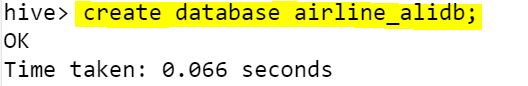


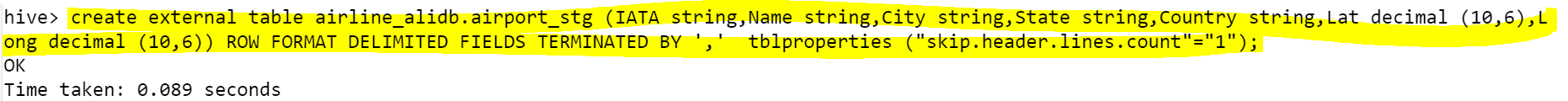


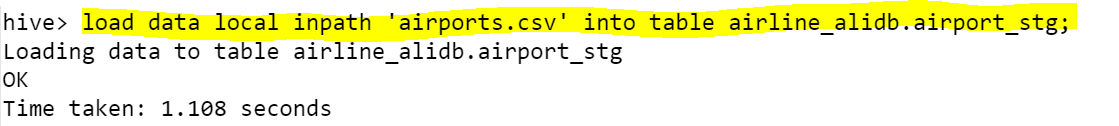


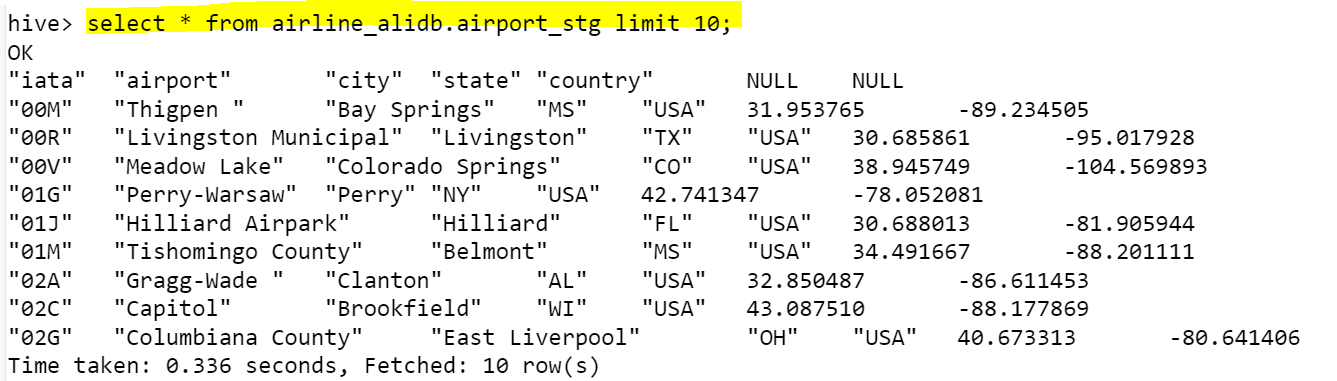
1. **Create tables using Hive.**

**Airport**

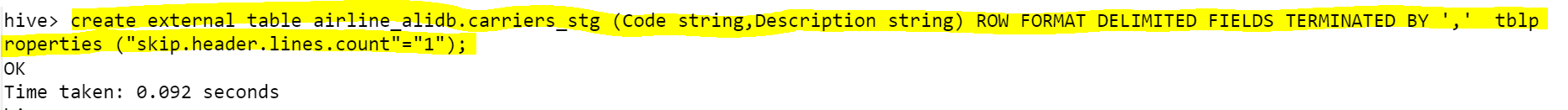


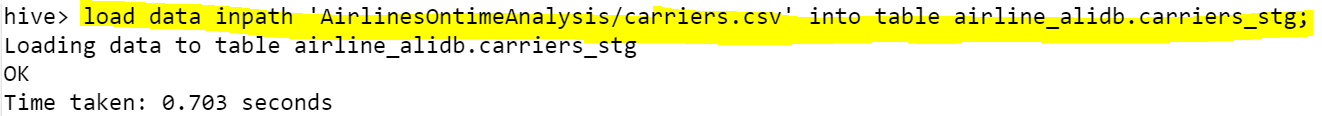






**Carriers**

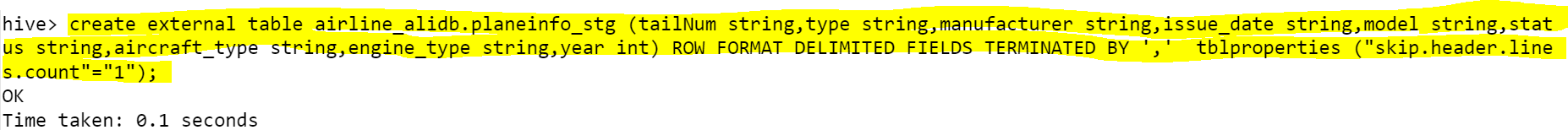


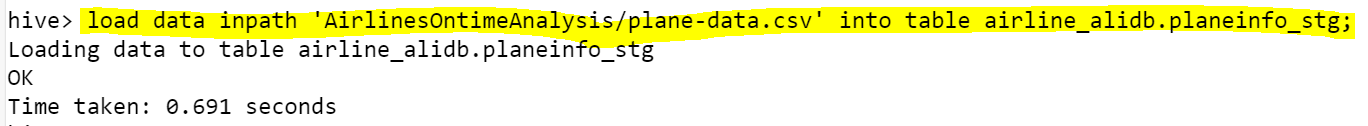


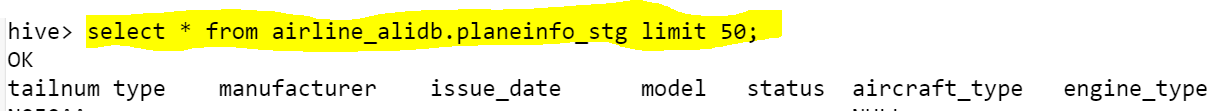
A screenshot of a computer

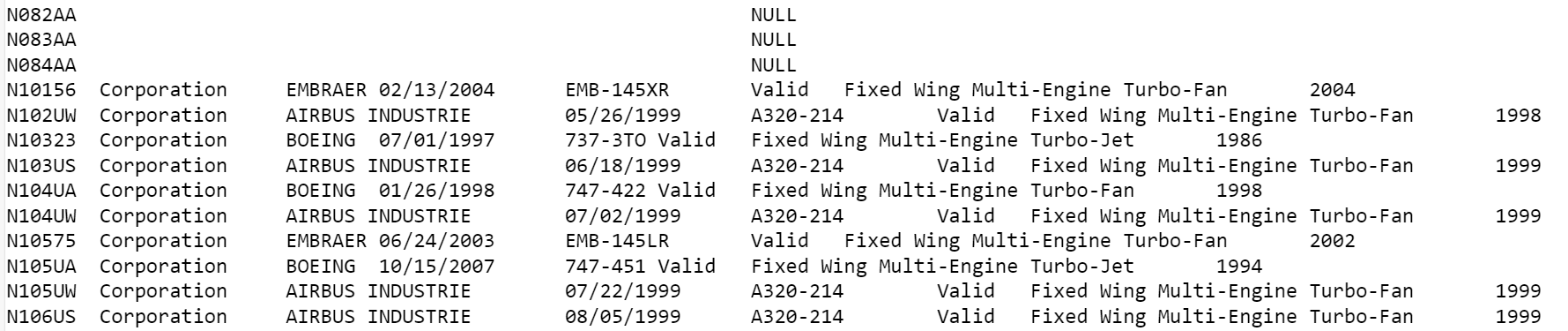
Description automatically generated

**Plane Info**

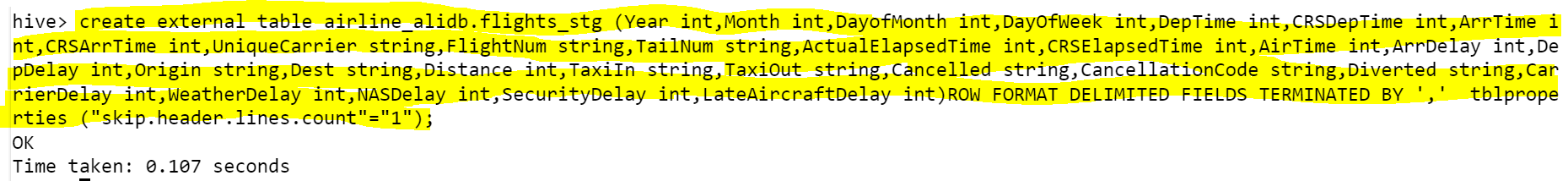


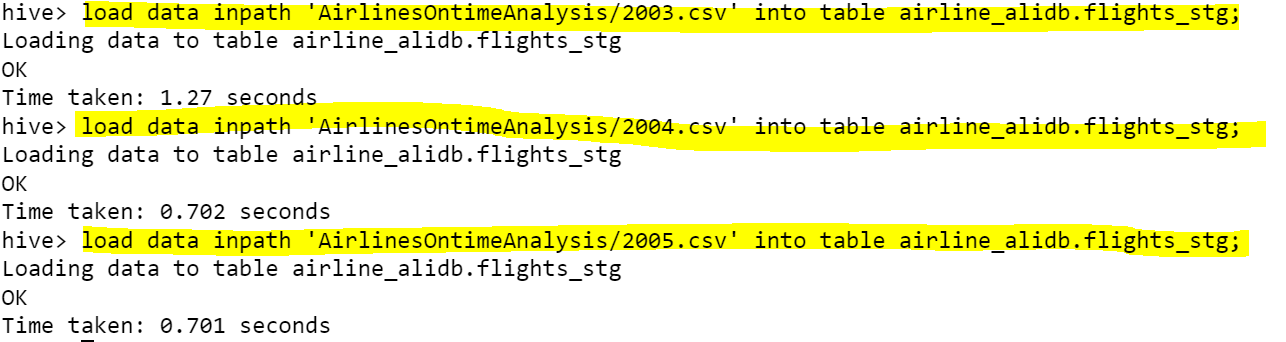


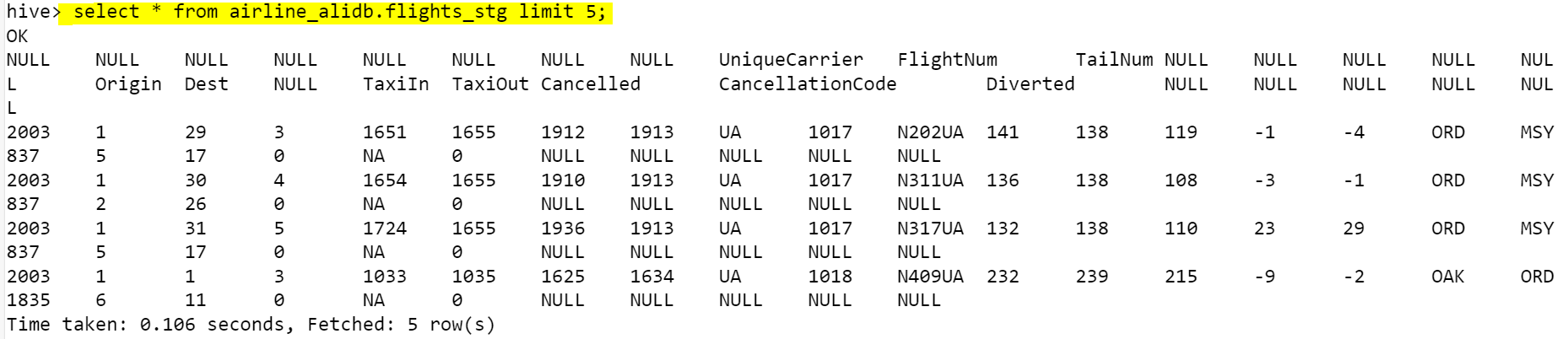




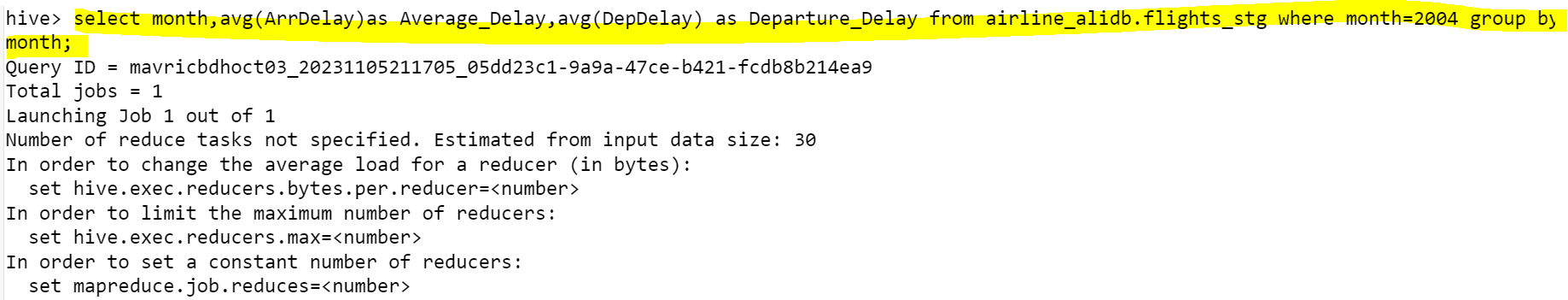
**Flights**







**What is the average arrival delay and average departure delay in each month of the year 2004?**



A close-up of a computer screen

Description automatically generated

1. **Improving the performance of Hive tables - Parquet format.**

